

多元时间序列的分割方法及拐点识别研究

摘要

绪论

时间序列分析发展过程

期货交易研究现状

第五章

优势

基于拐点分析的期货交易策略构造方法

基于目标时间序列数据提取IF-THEN关联规则

将IF-THEN规则转为TB交易平台上的交易策略

对交易策略进行评估，得到最优策略

多元时间序列的分割方法及拐点识别研究

摘要

1. 本文通过动态因子模型将多元时间序列变换为一元时间序列来适应已有的一元时间序列的分割算法从而达到分割多元时间序列的目的。

首先，通过动态因子模型从多元时间序列中抽取出一元公共因子序列，然后将二值分割算法、片段邻域算法以及修剪精确线性时间算法等三种典型的搜索算法用于公共因子序列的分割，该序列的分割结果被作为多元时间序列的分割结果。

2. 基于信息微粒和动态时间规整的构造商业周期的新方法。将GDP纳入考虑范围。

3. 期货价格数据具有波动性大、不易预测等特点。传统的时间序列线性模型，包括自回归(AR)模型以及自回归移动平均模型(ARMA)难以对期货价格数据进行有效预测。鉴于此，本文基于拐点分析结合布林带技术指标以及关联分析建立了一套用于期货自动交易的交易策略。

首先，基于期货价格数据构建相关的布林带指标，并应用模糊化方法提取布林带指标与未来价格变动指标问

的具有良好语义解释的关联规则关系。然后，在历史可以重演的假设下，提取的关联规则被设计成交易策略在期货交易软件交易开拓者上进行期货的程序化交易。最终，交易结果显示提取的规则能较准确地预判期货行情发生反转处的拐点，从而指导交易者在期货市场进行适当的交易来获益。

绪论

时间序列分析发展过程

时间序列分析的方法可以分为频域分析方法和时域分析方法这两类方法。频域分析方法主要包括谱分析和小波分析等，时域分析方法则包括自相关分析和交叉相关分析等。此外，时间序列分析方法也可以按照有参和无参分为两大类方法。有参方法假定所研究的平稳随机过程可以通过少量的参数来描述其确定的结构(例如，自回归模型或移动平均模型等)，这类方法的目的是估计出描述随机过程内部结构的参数。与之不同，无参方法并不假设随机过程有任何特殊结构而是直接估计随机过程的协方差和频谱等。时间序列分析方法也可以按照线性与非线性、一元时间序列与多元时间序列等标准来进行分类。

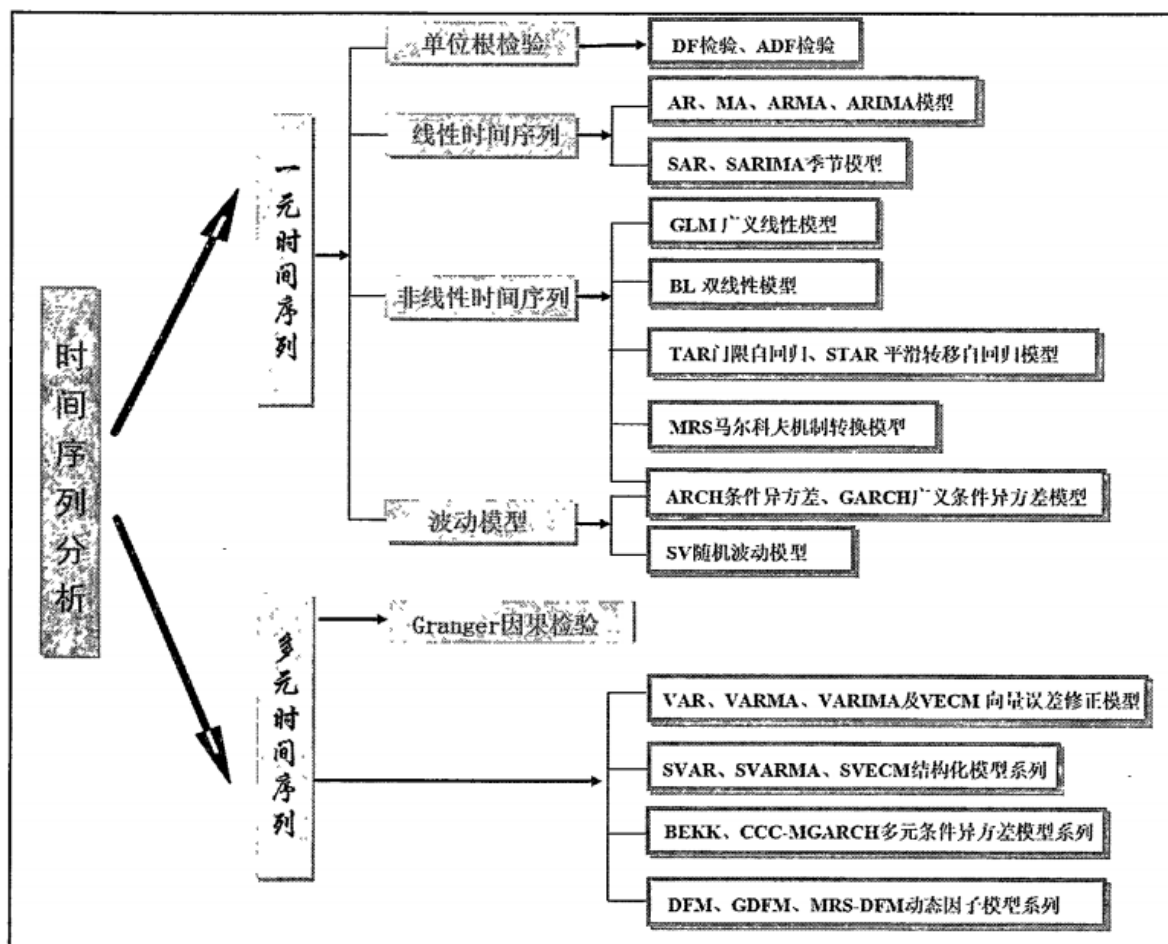


图 1.1 时间序列模型列表
Fig. 1.1 The list of time series models

Box和Jenkins所写的《Time Series Analysis : Forecasting and Control》给出了时间序列分析的一整套完整理论，包括数据的预处理、单位根检验以及模型的识别、估计、诊断和预测等。书中提出的**自回归积分移动平均模型**，又被称为Box—Jenkins模型，在处理具有明显趋势或具有季节性特征的时间序列时取得了良好效果。

Granger和Andersen在1978年提出了**双线性模型(Bilinear Model, BL)**来处理具有非线性特征的时间序列，这个模型可以看作是ARMA模型的直接推广，相较于ARMA模型它能更好地描述现实世界的复杂非线性现象。

Tong提出了基于分段线性思想的**门限自回归(Threshold Auto—Regressive, TAR)模型**来对非线性系统中的周期性数据建模，很好地提取到数据的非线性特征。

期货交易研究现状

期货市场的参与者按交易目的分类可以分为套期保值者和投机者，套期保值者是指前面提到的为了控制实际生产中的现货价格波动风险的生产商和加工商那类人，而投机者则是指的通过在期货市场上进行买卖操作实现以小博大获取利润的投资者。

第五章

金融市场的动态发展变化在时间上也表现为具有一定的连续性的过程。

两种类型的交易者：顺势交易者和逆向交易者。

这种逆向交易的思维其实是第四章中宏观经济周期中的拐点思想的一种体现，只是时间序列预测方法的滞后性决定了它无法及时侦测到拐点的出现，也就是行情的转变，因此经典时间序列的预测方法并不能很好的适用于交易市场。本章提出了一种基于拐点分析的期货交易策略的构建方法，这种方法不像经典的时间序列去预测未来的值，而是使用关联规则从期货价格的历史数据中提取出规律，判断交易行情中拐点出现的概率，从而指导交易者进行交易。

优势

1. IF—THEN形式的关联交易规则易于在期货交易软件交易开拓者(tradeblazer, TB)上进行算法实现;
2. IF—THEN形式的关联交易规则具有良好的语义解释，更符合人类的思维习惯，易于交易者理解和使用;
3. 计算得到的各类IF—THEN事件出现的概率可以为交易者进行交易提供客观依据，避免了传统预测方法的滞后性对交易带来的影响。

基于拐点分析的期货交易策略构造方法

基于目标时间序列数据提取IF-THEN关联规则

1. 对原始数据进行技术指标变换预处理

布林带

- 布林带中轨，长度为， z 个单位时间的滑动平均值(MA)，这里的单位时间可以是分、小时、天等不同的时间长度。
- 布林带上轨，即滑动平均值加上 r 倍的长度为 n 的时段内价格的标准差($MA + r\sigma$)
- 布林带下轨，即滑动平均值减去 r 倍的长度为 n 的时段内价格的标准差($MA - r\sigma$)

特点

- 布林线可以指示支撑和压力位置
- 布林带可以显示趋势
- 布林线具有通道作用
- 布林带可以显示超买和超卖
- 布林带可以给出止损的最佳点位

记号

上布林带指标: $u_t^{(n)} = \log\left(\frac{p_t}{\bar{p}_t^{(n)} + r \cdot v_t^{(n)}}\right)$

下布林带指标: $d_t^{(n)} = \log\left(\frac{p_t}{\bar{p}_t^{(n)} - r \cdot v_t^{(n)}}\right)$

价格差: $\Delta p_t = p_{t+1} - p_t$

数据选择

仅选取价格落在上布林带上方和下布林带下方的点(即 $u_t^{(n)} > 0$ 和 $d_t^{(n)} < 0$ 进行分析，将满足条件的 $u_t^{(n)}$ 以及与之对应时刻的价格差 Δp_t 作为目标数据，类似地，将满足条件的 $d_t^{(n)}$ 以及与之对应时刻的价格差 Δp_t 也作为目标数据。

2. 对数据进行模糊C均值(Fuzzy C-means, FCM)聚类

最大化划分到同一类中的对象的相似度，最小化划分到不同类的对象的相似度。

将样本个数为 N 的样本集 $Y = \{y_1, y_2, \dots, y_N\}$ 划分成 C 个模糊集，对应的类原型中心为 v_i , ($i = 1, 2, \dots, c$)，隶属度 u_{ij} 表示样本点 y_i 隶属于以 v_i 为聚类中心的模糊集的程度，对于 $\forall i, j: u_{ij} \in [0, 1]$ 且满足归一化条件：

$$\sum_{i=1}^c u_{ij} = 1, j = 1, 2, \dots, N.$$

记 $U = [u_{ij}] \in R^{c \times N}$ 为隶属度矩阵, $V = \{v_1, v_2, \dots, v_c\}$ 为类原型中心集合, 目标优化函数为:

$$J(U, V) = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{j=1}^N (u_{ij}^m d_{ij}^2) \right)$$

其中, m 为模糊化因子, 能影响聚类的效果, 满足 $m > 1$, 实际应用中通常取值为2。 d_{ij} 表示样本点 y_j 到聚类中心 v_i 的距离 $d_{ij} = \|y_j - v_i\|$, 通常用的是欧氏距离。记

$\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_N\}$ 为 N 个约束式拉格朗日乘子集合, 则, 则该目标优化函数在归一化约束条件下可以转化为如下的拉格朗日算子函数:

$$\begin{aligned} L(U, V, \Lambda) &= J(U, V) + \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) \\ &= \sum_{i=1}^c \left(\sum_{j=1}^N (u_{ij}^m d_{ij}^2) \right) + \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) \\ &= \sum_{i=1}^c \left(\sum_{j=1}^N (u_{ij}^m \|y_j - v_i\|^2) \right) + \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) \end{aligned}$$

关于参数 u_{ij} 和 $c_i (i = 1, 2, \dots, c, j = 1, 2, \dots, N)$ 求导并令其等于0得到表达式:

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{\|y_j - v_i\|}{\|y_j - v_k\|} \right)^{2/(m-1)} \right]^{-1} \quad (5.9)$$

$$v_i = \frac{\sum_{j=1}^N (u_{ij}^m y_j)}{\sum_{j=1}^N u_{ij}^m} \quad (5.10)$$

FCM聚类算法流程:

1. 选择类原型的数量 c , 模糊化因子 m 和终止条件 ϵ 的值, 选取距离函数(通常是Euclidean距离), 初始化划分矩阵 U^0 ;
2. 当迭代次数为 $l, (l = 1, 2, \dots)$ 时, 根据 U^{l-1} , 利用式(5.10)更新聚类原型 V^l ;
3. 根据 V^l , 利用式(5.9)更新划分矩阵 U^l ;
4. 如果对于 $\forall i, j: \max |u_{ij}^l - u_{ij}^{l-1}| < \epsilon$, 算法停止, 返回划分矩阵 U 和类原型 V ; 否则, $l = l + 1$, 返回步骤(2)。

运用FCM聚类算法对目标数据每一个属性上的数据进行聚类, 并得到对应的类原型中心。

3. 挖掘关联规则

数据之间经常会存在相关关系, 但是又往往难以用具体的函数关系去刻画这种相关关系。关联规则方法基于数据统计的方法, 通过频数统计、支持度、置信度计算等方法能有效表示数据之间的关联关系。上、下布林带指标以及价格差指标都是连续型数据并不适合做关联分析, 因此考虑将数据离散化来进行关联规则挖掘。

隶属函数是用来表示某元素与该元素属于某个模糊集的程度(也称为隶属度)的对应关系, 隶属函数的值域范围是 $[0, 1]$ 。

将IF. THEN规则转为TB交易平台上的交易策略

关联规则前件 $u_t^{(n)}$ 属于 $FS_{u,i}$ 作为交易策略的入场条件, 即 $u_t^{(n)}$ 达到一定值的范围时入场建仓, 至于是建多仓(买入)还是空仓(卖出)取决于后件 Δp_t 属于 $FS_{\Delta,j}$ 的情况, 如果 $\Delta p_t > 0$ 表示价格会上涨, 则买入建多仓; 如果 $\Delta p_t < 0$, 表示价格会下跌, 则卖出建空仓。

对交易策略进行评估, 得到最优策略